

AD-A269 017



ION PAGE

Form Approved

OMB No. 0704-0188

Public
Sale
Collection
Davis Highway, Suite 404, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

age 1 hour per response, including the time for reviewing instructions, searching existing data sources, collection of information, send comments regarding this burden estimate or any other aspect of this Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 31, 1993	3. REPORT TYPE AND DATES COVERED Final	
4. TITLE AND SUBTITLE Validity of Integrity Tests for Predicting Drug and Alcohol Abuse			5. FUNDING NUMBERS C No. N00014-92-J-4068	
6. AUTHOR(S) Frank L. Schmidt Chockingam Viswesvaran Deniz S. Ones				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Iowa Iowa City, IA 52242			8. PERFORMING ORGANIZATION REPORT NUMBER GB22 612025110-331	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Defense Personnel Security Research Center Monterey, CA 93940-2481			10. SPONSORING/MONITORING AGENCY REPORT NUMBER DTIC ELECTE SEP 07 1993 S A D	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT <div style="border: 1px solid black; padding: 5px; display: inline-block;">This document has been approved for public release and sale; its distribution is unlimited.</div>			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This research used psychometric meta-analysis (Hunter & Schmidt, 1990b) to examine the validity of integrity tests for predicting drug and alcohol abuse. For both drugs and alcohol, integrity test scores correlated substantially (.31 to .51) with admissions of abuse in student and employee samples. In samples of job applicants, however, the mean validity was lower (.21) for drug abuse; for alcohol abuse validity for applicants was high but only one study (N = 320) was found. All meta-analysis indicated that validity was generalizable. We conclude that the operational validity of integrity tests for predicting drug and alcohol abuse in the workplace is probably about .30.				
14. SUBJECT TERMS Drug Abuse, Alcohol Abuse, Security Screening, Personality, Integrity Testing			15. NUMBER OF PAGES 94	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

Validity of Integrity Tests for Predicting
Drug and Alcohol Abuse

Frank L. Schmidt
Department of Management and Organizations
University of Iowa

Chockalingam Viswesvaran
Department of Psychology
Florida International University

Deniz S. Ones
Department of Management
University of Houston

Submitted to: Program Clearance Processes,
Personnel Security Research Award Program,
PERSEREC, Office of Naval Research, Monterey, CA

August, 1993

In Fulfillment of: Contract No. N00014-92-J-4068
Institution: University of Iowa

93-20619
9728

93 9 02 077

ACKNOWLEDGMENTS

Preparation of this technical report was supported by the Office of Naval Research (Contract No: N00014-92-J-4068 to Frank Schmidt). We would like to thank Howard Timm and Ralph Carney of the Defense Personnel Security Research and Education Center for their encouragement and support of this research. The content of this technical report does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

We acknowledge with thanks the contributions to this research of numerous test publishers, authors, and colleagues. We also acknowledge with gratitude the test publishers and researchers for sharing their research for this study. We would also like to express our gratitude to Marcia Simmering, Sara Schultz, and Michelle Davis for their help in typing and data management.

Parts of this technical report were presented in the 1993 annual conference of the American Psychological Association, Toronto, Canada, and in the 1992 annual conference of the American Psychological Association, Washington, D.C..

Integrity and Substance Abuse

Correspondence regarding this technical report should³
be addressed to Frank L. Schmidt, Department of Management,
University of Iowa, Iowa City, IA 52242.

Accession For		/
NTIS	CRA&I	<input checked="" type="checkbox"/>
DTIC	TAB	<input type="checkbox"/>
Unannounced		<input type="checkbox"/>
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and/or Special	
A-1		

DTIC COLLECTION

ABSTRACT

Integrity tests have previously been found to predict other counterproductive workplace behaviors (e.g., absenteeism, property damage, and violence on the job; Ones et al., in press). This research used psychometric meta-analysis (Hunter & Schmidt, 1990b) to examine the validity of integrity tests for predicting drug and alcohol abuse. All studies included in this meta-analysis were concurrent in nature. For both drugs and alcohol, integrity test scores correlated substantially (.31 to .51) with admissions of abuse in student and employee samples. In samples of job applicants, however, the mean validity was lower (.21) for drug abuse; for alcohol abuse validity for applicants was high but only one study (N = 320) was found. All meta-analysis indicated that validity was generalizable. Based on our analyses, we conclude that the operational validity of integrity tests for predicting drug and alcohol abuse in the workplace is probably about .30. But further research is needed; predictive validity studies conducted on applicants are particularly needed.

EXECUTIVE SUMMARY**STATEMENT OF THE PROBLEM:**

Drug and alcohol abuse is a major problem in the workplace. In this report, we investigate the validity of paper and pencil measures of integrity for predicting substance abuse. In environments that require high levels of security, paper and pencil measures assessing integrity can be useful for screening of job applicants. To the extent that selection methods can be used to eliminate substance abusers at the point of hire, drug testing programs for employees become less necessary. The less obtrusive nature of integrity tests compared to drug tests makes them attractive for screening purposes. The validity of integrity tests for substance abuse can be used in evaluating relative advantages over other alternative methods of screening for drug and alcohol abuse. Thus, this research can also aid in the development of new, and more effective instruments for personnel screening.

Further, we also examine the moderating influences on the validity of integrity tests for predicting substance abuse. Specifically, we wanted to examine the following potential moderators of validity of integrity tests in predicting substance (alcohol and drug) abuse:

1. Type of test (overt vs. personality-based tests)

2. Type of scale (drug vs. other scales)
3. Criteria based on self report vs. criteria based on external measurement
4. Predictive vs. concurrent validity studies
5. Validation sample (applicants vs. employees vs. students)
6. Job complexity

METHODS AND DATABASE USED:

A comprehensive search of published and unpublished literature resulted in the location of 50 validation studies involving 25,594 individuals. Psychometric meta-analysis (Hunter & Schmidt, 1990b) was used to correct for errors and biases in the individual studies, and cumulate the results across the 50 studies. Of these fifty studies, 24 had used employees as samples, 16 had used student samples, and the remaining 10 studies were based on applicant samples. All fifty studies employed the concurrent validation strategy. Forty-eight of the fifty studies had relied on admissions (self-reports) of substance abuse. There was one study conducted in a sample of 46 employees in a fire department that had used apprehension and conviction for substance abuse as the criterion. The observed validity coefficient in that study was .44. One study provided inadequate information as to

whether admissions or external measures were employed. The observed validity coefficient in that study was .62 and it was based on a sample of 320 job applicants.

The admissions criterion was measured using self-report questionnaires. Measures of admissions of drug abuse included questions on number and type of illegal drugs used, number of times one has become "high" from drug use, etc. Measures of admissions of alcohol abuse included questions on frequency of alcohol intoxication, number of drinks consumed on the job, number of drinks on work breaks and during lunch on workdays, number of alcohol-related problems, etc. The final score was the sum (sometimes weighted) of such admissions.

Twenty of the fifty studies were conducted in the mid west while four were conducted in the north western region of the United States. Thirteen of the fifty studies were conducted in supermarket or grocery stores or convenience stores or gas station employees. Seven of the fifty studies were done using security personnel as sample. One study was conducted in a fire department while another was in a fast food chain. Twenty studies focused on alcohol consumption while the remaining thirty used drug abuse as the criterion.

RESULTS:

Across 50 studies, the true validity of integrity tests for predicting substance abuse (drug and alcohol abuse combined) was .26. The standard deviation of the true score validity was .14 across the 50 studies. This value is small in relation to comparable figures from other predictor domains. The 90% credibility value was .10. That is, 90% of the estimated true validities are higher than .10.

The separate true validities for student, employee, and applicant populations for combined drug and alcohol abuse were .48, .36, and .22, respectively. It is of interest to note that most of the sample consisted of applicants (about 90%). This is significant because in a selection setting, the focal population of interest is the applicant population. Many researchers have argued (see Ones et al., in press, for a summary) that conscious and/or unconscious response distortion will affect integrity test validities. In taking these tests, applicants have the greatest incentive for response distortion, followed by employees and students, in that order. That is, to the extent integrity test validities are affected by response distortion, true validities based on applicant samples should be lower than true validities based on employee

samples, which in turn should be lower than the true validities computed on student samples.

The results of our analyses confirm this expected gradient. But, although response distortion on the tests seems to attenuate the validity of integrity tests, its effects do not destroy predictive validity. Even in the applicant population the true validity was .22 and the 90% credibility value was .13. Although this level of validity is moderate, these values suggest that the use of integrity tests in employment selection will translate into reduced levels of substance abuse in the workplace.

But response distortion on the predictor side is only part of the problem when (a) the criterion used for validation is admissions of substance abuse; and (b) concurrent validation strategy is employed. Response distortion could occur on the criterion measure when the criterion used is admissions of substance abuse. Response distortion on the predictor [test] does not bias estimates of operational predictive validity, because it reflects the reality that will hold when the test is used in hiring applicants. That is, real applicants will display some response distortion. Response distortion on the criterion, on the other hand, will bias predictive validity downward. Further, all validities in this meta-analyses were

concurrent. The criterion for applicants was admissions of drug abuse made at the time they were applicants. Use of this same criterion measure taken later after the participants had been on the job for some time would have given a better indication of predictive validity. Because in predictive studies there may be less response distortion on the (admissions) criterion measure, predictive validity estimates might be higher than the .22 reported here.

Specifically, with admissions as the criterion measure, concurrent studies done on applicants may underestimate predictive validity computed on applicants. Concurrent studies done on applicants using admissions will strongly lend themselves to response distortion on the criterion measure, which in turn would bias validity estimate downward. Applicants for jobs have strong incentives to minimize admissions of previous illegal drug use. Present employees already have jobs, and in addition are usually told that their responses will be used for research purposes only. So present employees have much less incentive for response distortion on the criterion.

Given these biases, the actual operational validity of integrity tests for predicting drug abuse is probably somewhere between the validity of .22 (estimated with applicant samples) and .36 (obtained from employee

samples). This value is large enough to produce practically significant reductions in substance abuse on the job if integrity tests are used in hiring.

CONCLUSIONS AND RECOMMENDATIONS:

- Integrity test validities are substantial and generalize across situations. Use of integrity tests will result in substantial utility gains.
- More primary studies with different designs (e.g., predictive validation) and jobs of varied complexity need to be done. This will facilitate a more comprehensive fully hierarchical meta-analysis in the future.
- Primary research studies should report reliabilities (especially for the criterion measures) and range restriction.
- We found a gradient in true validity across student, employee and applicant samples (true validity was highest in student samples). Future research should test the effects of faking and conscious dissimulation on predictive validity.

- Future research should explicitly test the three causal mechanisms (hypothesized in this report) that explains the validity of integrity tests for predicting substance abuse.

CHAPTER I

THE PROBLEM OF SUBSTANCE ABUSE

Substance abuse is a major societal problem. Numerous surveys (e.g., Johnston, O'Malley, & Bachman, 1986; Miller et al., 1983) have found that substance abuse, especially the consumption of alcohol and marijuana, is prevalent in the general population. Epidemiological surveys (e.g., Simpson, Curtis, & Butler, 1975) indicate that substance abusers are predominantly in the age group 21-25 years and mostly male.

The relationship between substance abuse and job performance and other job related behaviors has been studied. McDaniel (1988) found in a large sample study of military personnel that individuals who reported using drugs at earlier ages were more likely to be rated as unsuitable for service by their supervisors than a control group who did not use drugs when younger. In a sample of Navy recruiters, Blank and Fenton (1989) found that individuals testing positive for drugs had more behavioral and performance problems than individuals who tested negative for drugs.

Normand, Salyards, and Mahony (1990) found that postal employees who tested positive for substance abuse were more

likely to be absent from work. Further, Winkler and Sheridan (1989) found that employees who entered employee assistance programs for treating drug addiction were more likely to be absent, had twice the number of worker compensation claims, and used more than twice as many medical benefits as a matched control group. Crouch, Webb, Peterson, Butler, and Rollins (1989) found that drug use correlated with increased accident and absence rates.

Substance abuse has been found to be related not only to measures such as absenteeism, turnover, accidents, and productivity, but also to related behaviors such as stealing on the job, violence, and effort expenditure (i.e., not daydreaming) on the job. In fact, Viswesvaran (1993) found that all these various measures of job performance are positively correlated and that a general factor exists across the different measures, suggesting that the various measures of job performance may be caused in part by the same underlying construct (presumably a personality dimension). That is, a hierarchical model involving a general factor explained the true score correlations between the different measures of job performance indicating that the various measures of job performance could be construed as manifestations of the same underlying construct.

In addition to the above mentioned studies that compare individuals using drugs to a matched set of controls on various job performance measures, several laboratory studies have also found that substance abuse leads to impairment in performance of various experimental tasks (e.g., Herning, Glover, Koepfel, & Jaffe, 1989; Jobs, 1989; Streufert et al., 1991; Yesavage, Leirer, Denari, & Holister, 1985). Impairment in information processing capabilities, decision making, slowing of reflexes have been found to result from drug or alcohol consumption.

In summary, surveys indicate that substance abuse is prevalent in the general population, and studies show a negative relationship between substance abuse and job performance. This suggests that employers, co-workers, customers, and the general public all have a stake in reducing drug and alcohol use in the work place. Employers have tried different strategies to ensure a drug free work place.

Employee drug testing has increased over the past few years (Freudenheim, 1988). The increasing concern of organizations with drug-related issues is justified by the negative effect drug abuse has on the organization's bottom line. Drug abuse, as indicated earlier, has been linked to a variety of organizational costs, including accidents,

lost productivity, and health care (Berry & Boland, 1977; Konovsky & Cropanzano, 1991; Trice & Roman, 1972). This increasing concern of employers with drug abuse has resulted in increased drug testing of both current and prospective employees for drug abuse (Guthrie & Olian, 1986).

A survey of the literature indicates that employer strategies are mainly based on four considerations: (a) the validity and reliability of the techniques used to detect substance abuse; (b) the legal viability of the techniques; (c) the practicality of employing the techniques (i.e., is it feasible to use that technique; obviously, the employer cannot place all employees under surveillance round the clock); and (d) whether employees accept the use of that technique as justified.

Validity refers to whether the technique is measuring what it purports to measure. Reliability indicates whether the measurements are replicable (and not due to some extraneous element at the time the measurement is made). Legal viability refers to the employers' concerns about whether the courts and arbitrators will accept the findings of the technique. In fact, studies have shown (see summary in Hill & Sinicropi [1987]) that courts and arbitrators place considerable weight on the reliability and validity

of the technique used in deciding cases involving substance abuse. Thus, the validity and reliability of the technique has an indirect effect on the strategies used by the employers to combat substance abuse, as well as a direct effect.

Employee acceptability of drug testing programs has been widely researched. Negative employee reactions to drug testing, if ignored, may lead to lowered commitment and subsequent reduction in performance (Crouch et al., 1989). Konovsky and Cropanzano (1991) present data indicating that employee reactions to drug testing can be analyzed within an organizational justice framework (Adams, 1965; Greenberg, 1990). Specifically, Konovsky and Cropanzano (1991) found that perceptions of procedural justice affect reactions to drug testing. Two of the key elements in shaping perceptions of procedural justice are: (a) the validity, reliability, and psychometric properties of the testing procedures; and (b) invasions of privacy concerns. Other elements include job characteristics (i.e., people accept drug testing when impaired performance results in dangers to others; see Stone & Vine, 1989); type of drug used (Murphy, Thornton, & Reynolds, 1990); the personnel action taken against employees testing positive (Gomez-Mejia & Balkin, 1987; Stone & Kotch, 1989); the role of explanations (Bies, 1987; Bies & Shapiro, 1987; Crant &

Bateman, 1989); the chance to appeal; the availability of¹⁸ advance notice; and whether random testing or testing with due cause is implemented. Employee objections could result in union contracts restricting the use of certain techniques of detecting substance abuse. Further, courts and arbitrators are likely to give some weight to employee and applicant objections in their decisions. Thus, employee acceptability has both a direct effect and an indirect effect (through legal acceptability) on the strategies used by an employer.

However, surveys also indicate a distinction in acceptability reactions depending on whether drug testing is intended for applicants or employees. In fact, surveys (e.g., Stecher & Rosse, 1992) indicate that drug testing for selection evokes less antagonism than drug testing of satisfactorily performing employees. Both the applicants and the general public (including employees, unions, arbitrators, and courts) are more receptive of drug testing during hiring than drug testing of current employees (when the employer is expected to provide a just cause for testing). For approving drug testing of applicants, the single most important issue seems to be the validity and reliability of the instrument used.

In short, the validity and reliability of the instrument used affects legal defensability of the procedures, acceptability to test takers, as well as directly affecting the employers' choice of technique used. Further, the validity and reliability of the technique affects the strategies used by the employer through its an effect on legal defensability and acceptability to test takers. The important role (both direct and indirect) played by the validity and reliability in the choice of the techniques is pictorially depicted in Figure 1. Thus, it is of paramount interest to examine the validity and psychometric properties of the procedures used for drug testing to realize the benefits of drug testing without loss of employee commitment.

Insert Figure 1 about here

Several approaches have been tried to detect drug abuse. Blood testing, breathe analyzers, urinalysis are some of the common approaches to drug testing and detection. One technique that is gaining prominence in employment settings is the use of paper and pencil pre-employment integrity tests to assess a job applicant's

predisposition to drug and alcohol abuse. Evidence available to date indicates that applicants do not object to such tests (Stecher & Rosse, 1992; Stone & Bommer, 1990; Stone & Kotch, 1989). Further, integrity tests are paper and pencil measures and are not physically intrusive. To the extent that selection methods can be used to eliminate drug abusers at the point of hire, drug testing programs for employees become less necessary. In the next chapter, we discuss the theoretical underpinnings of these tests that could explain their validity for predicting substance (drug or alcohol) abuse.

CHAPTER II

INTEGRITY TESTS AND SUBSTANCE ABUSE

This chapter is organized as follows. We first define what we include as integrity tests. Following this definition, we present a brief history of integrity tests and their development to the current stage. Then we review the literature that examines the personality constructs underlying integrity tests. Finally, we discuss some causal mechanisms hypothesized in the literature by which the personality constructs assessed by integrity tests can predict substance abuse. That is, we first identify the personality constructs tapped into by integrity tests; then discuss the theoretical and conceptual basis by which the personality constructs assessed by integrity tests could be related to substance abuse.

Integrity Tests

Defining Integrity Tests

Integrity tests are paper and pencil measures designed to measure the predispositions of individuals to engage in counterproductive behaviors on the job. Integrity tests are paper and pencil tests, as opposed to other methods such as the polygraph (a physiological method), background

investigations, interviews, and reference checks. These tests have been developed for use with applicants and employees (a normal population); hence instruments such as the MMPI, which were designed for use with mentally ill population, are not classified as integrity tests even though some organizations claim to use them for screening out delinquent applicants (see Ones, 1993, for further elaboration of the characteristics of integrity tests). Most integrity tests have been initially designed to predict a variety of counterproductive behaviors; only later were they found to predict other criteria such as supervisory ratings of overall job performance (Ones, Viswesvaran, & Schmidt, 1993).

A Brief History of Integrity Tests

The first paper and pencil psychological test to assess the integrity of potential employees, the Personnel Reaction Blank, was developed in 1948 (Gough, 1948). It was a derivative of what was then called the Delinquency scale of the California Psychological Inventory. (This scale was later renamed the Socialization scale.) In 1952, a second type of test, intended to assess honesty of job applicants, was developed. This test, the Reid Report, was a compilation of questions that seemed to distinguish honest and dishonest individuals during polygraph

examinations. Since then several other instruments have been developed and used to select applicants on the basis of integrity. A complete treatise of the history on integrity tests can be found in Ash (1989) and Woolley (1991).

There is relatively little information about companies that use paper and pencil integrity tests. According to Sackett and Harris (1985) as many as 5,000 companies may use pre-employment integrity tests, assessing about 5,000,000 applicants yearly. A variety of surveys of companies indicate that anywhere between 7 to 20% of all companies in the US could be using integrity tests in hiring for at least for some jobs. For various estimates see American Society for Personnel Administration, 1988; Blocklyn, 1988; Bureau of National Affairs, Inc., 1988; O'Bannon, Goldinger, & Appleby, 1989. Even by the most conservative estimates, millions of people in the US have been tested using integrity tests. There are at least 43 integrity tests in current use. Ones (1993) observes that of these tests, about a quarter seem to be small operations without much market share and overall 16-19 tests seem to serve the majority of the demand for integrity tests. However, this demand may be increasing because in 1988 the Federal Polygraph Act effectively banned the use of the polygraph in employment settings.

Employers' desire for trustworthy and conscientious employees has spawned a multimillion dollar industry of integrity testing (see O'Bannon et al., 1989 for prices of various integrity tests three years ago). Employers' concern regarding counterproductive behaviors at work coupled with the recent passage of the Employee Polygraph Protection Act (1988) seems to indicate that paper and pencil integrity tests will be more broadly used in the future than they are today.

Over the last fifteen years, scientific interest in integrity testing has increased substantially. The publication of a series of literature reviews attests to the interest in this area and its dynamic nature (Guastello & Rieke, 1991; Sackett, Burris, & Callahan, 1989; Sackett & Decker, 1979; Sackett & Harris, 1984). Recently Sackett et al. (1989) and O'Bannon et al. (1989) have provided extensive qualitative reviews and critical observations regarding integrity testing. In addition to these reviews, the US Congressional Office of Technology Assessment (OTA) (1990) and the American Psychological Association (APA) (Goldberg, Grenier, Guion, Sechrest, & Wing, 1991) have each released "papers" on integrity tests. The OTA paper (1990) was in part prompted by the Congress' regulation of the polygraph. The OTA recommendations were based on the results of only a few "technically competent" studies,

ignoring most of the literature on integrity tests.

Compared to the OTA paper (1990), the APA report (Goldberg et al., 1991) was more thorough, objective, and insightful. It provided a generally favorable conclusion regarding the use of paper and pencil integrity tests in personnel selection.

Personality Constructs Underlying Integrity Tests

Sackett et al. (1989) classify honesty tests into two categories: "Overt integrity tests" and "Personality-based tests." Overt integrity tests (also known as clear purpose tests) are designed to directly assess attitudes regarding dishonest behaviors. Some overt tests specifically ask about past illegal and dishonest activities as well; although for several admissions are not a part of the instrument, but instead are used as the criterion. Overt integrity tests include the London House Personnel Selection Inventory (PSI) (London House, Inc., 1975), Employee Attitude Inventory (EAI) (London House, Inc., 1982), Stanton Survey (Klump, 1964), Reid Report (Reid Psychological Systems, 1951), Phase II Profile (Lousig-Nont, 1987), Milby Profile (Miller & Bradley, 1975), and Trustworthiness Attitude Survey (Cormack & Strand, 1970). According to Sackett et al. (1989), " . . . the underpinnings of all these tests are very similar . . . "

(p. 493). Hence, high correlations may be predicted among²⁶
all these overt integrity measures.

On the other hand, personality-based measures (also referred to as disguised purpose tests) aim to predict a broad range of counterproductive behaviors at work (e.g., violence on the job, absenteeism, tardiness, drug abuse, in addition to theft) via personality traits, such as reliability, conscientiousness, adjustment, trustworthiness, and sociability. In other words, these measures have not been developed solely to predict theft or theft-related behaviors. Examples of personality-based measures that have been used in integrity testing include the Personal Outlook Inventory (Science Research Associates, 1983), the Personnel Reaction Blank (Gough, 1954), Employment Inventory of Personnel Decisions Inc. (Paajanen, 1985), and the Hogan's Reliability Scale (Hogan, 1981). Different test publishers claim that their integrity tests measure different constructs, including responsibility, long term job commitment, consistency, proneness to violence, moral reasoning, hostility, work ethics, dependability, depression, and energy level (O'Bannon et al., 1989). The similarity of integrity measures raises the question of whether they all measure primarily a single general construct. Detailed descriptions of all the above tests can be found in the

10th Measurement Yearbook (Conoley & Kramer, 1989) and in the extensive reviews of this literature (O'Bannon et al., 1989; Sackett et al., 1989; Sackett & Harris, 1984).

Many factor analytic investigations have been conducted on a number of integrity tests. More factor analytic investigations have been conducted on overt integrity tests than on personality-based integrity tests. Cunningham and Ash (1989) investigated the dimensionality of the Reid Report using principal components analysis using two large samples (N's of 1,281 and 3,071). They found that a solution of four interpretable factors fit the data best (the four factors were labeled self punitiveness, punitiveness toward others, self projection, projection toward others). Jones and Terris (1984) examined the factor structure of the PSI and found six factors (these were labeled theft temptation and rumination, theft rationalization, projection of theft in others, theft punitiveness, inter-thief loyalty, personal theft admissions). Harris and Sackett (1987) also investigated the factor structure of the PSI Honesty scale (N=349 job applicants) and found four interpretable factors, which they labeled temptation and thoughts about dishonest behaviors, actual and expected dishonest activities, norms about the dishonest behaviors of others, impulse control and behavioral tendencies. Martelli (1988) conducted a

principal components analysis of the Phase II Profile and found three factors. Hay (1981) and Harris (1987) investigated the factor structure of the Stanton Survey and found seven interpretable factors (these were labeled general theft, opportunism, employee theft, leniency, employee discounting, perceived pervasiveness of dishonesty, and association with dishonest individuals). However, both the attitudes and admissions part of the Stanton Survey were used, a decision that probably clouds the comparison of Stanton Survey factor structure with other overt tests.

A major shortcoming of these factor analytic studies is that no general factorial solution was investigated. In all of these studies, the investigators have aimed to confirm a multiple factorial model of integrity. In other words, factor analysts of integrity tests have never looked for a general factor. This is a major shortcoming. In fact, the multiple factors these researchers claim to have found are highly correlated, indicating a problem of overfactoring. This might also be intuitively evident from the labels different researchers used to describe the multiple dimensions (for example, in one study, general theft and employee theft were claimed to be separate dimensions). The results of different factor analytic studies reflect interpretations of various researchers, yet

there seems to be a degree of overlap in the constructs integrity test tap into. The assertion that overt integrity tests appear to be multidimensional does not preclude the establishment of a general factor. This interpretation is strengthened by a finding in many of the previously reviewed factor analytic studies. A first factor accounted for a large proportion of the variance when compared to subsequent factors. This fact coupled with high intercorrelations among factors clearly points to the presence of a general factor. Harris and Sackett (1987) explicitly stated that a general factor accounted for most of the variance in their data and further conducted Item Response Theory (IRT) analyses using the one parameter Rasch model. Their results suggested that the PSI Dishonesty scale taps into "an underlying construct which may be called dishonesty" (p. 134).

Relatively few studies have investigated the factor structure of personality-based integrity tests. Paaanen (1987) factor analyzed the PDI Employment Inventory. The PDI Employment Inventory has three scales: Performance, Tenure, and Frankness. Of these three scales, only Performance scale is considered to be a personality-based integrity test (even though the observed correlations between the Performance scale and the Tenure scale range between .45-.65). In Paaanen's factor analysis of the PDI

Employment Inventory (all three scales combined), a five factor solution provided the best fit to the data. These factors were labeled irresponsibility, sensation seeking, unstable upbringing, frankness and conforming motivation. Similar to the results for overt integrity tests, positive correlations were reported among the dimensions and a large proportion of the variance was accounted for by the first factor "irresponsibility," strengthening an argument for a general factor.

Moreover, most of these studies have examined the factor structure of individual integrity tests. Such studies are necessary and useful for refining lines of construct validity evidence for single instruments, but they are less useful when the focus is on investigating construct validity across measures. In addition, the proprietary nature of scoring keys for most integrity tests makes it impossible to factor analyze them. Positive and often fairly respectable correlations among group factors detected in factor analytic studies appears to be evidence of a general factor and further justifies the need to examine whether a general factor exists across measures.

Recently, Ones (1993) examined whether a general factor exists across tests. Using both primary data ($N = 1,365$) and meta-analytic cumulation, she found that a

general factor exists across different integrity tests. This finding is important because now researchers can focus on the theoretical construct underlying the different measures rather than investigating each measure separately as if each measure is unique. All theoretical propositions and causal explanations are stated in terms of constructs and not measures (Nunnally, 1978).

Ones (1993) also examined the correlation between composites of integrity test scores (a linear composite across different tests) and measures of personality dimensions (again a linear composite of different measures of the same construct). The objective in forming the composite was to define the general factor as what is common across all measures, which will be a more construct valid measure of that construct than any single measure that makes up the composite.

Jensen (1980, p. 223) uses measurement of height as an analogy to explain how the composite measure is a more construct valid measure. Consider the physical stature (height). Imagine a situation where we cannot measure an individual's height directly but can measure only the lengths of (a) lower leg, (b) upper leg, (c) torso, (d) neck, and (e) head. If these measurements could be made only on interval and not absolute (ratio) scales, we could

only express the standing of individuals on each of the five measures as a standard score. Now, if we were able to measure the height of the individuals directly on a true scale, we would find that the composite of the five measures correlates higher than any one of the five measures with the individual's total height measured in the true scale. That is, the composite is a more construct valid measure of the height of the individual than any one of the five measures of height considered separately. In forming the linear composite, we can use unit weights or weight the measures by their loadings on the general factor. Both composites will be more construct valid than the individual measures, and the difference between the two linear composites (unit weights vs. weighting by the general factor) in most cases will be small (Harman, 1976).

Ones (1993) found that the variance common to all integrity tests correlated highest with the personality dimension of conscientiousness, followed by agreeableness and then emotional stability (neuroticism). Based on her comprehensive analyses, we can conclude that integrity tests tap into the personality dimensions of conscientiousness, agreeableness, and emotional stability, in that order.

Review of Causal Mechanisms: Why Personality Constructs
Underlying Integrity Tests Might Predict Substance Abuse

Three causal mechanisms have been proposed in the literature that explains why personality constructs tapped into by integrity tests should predict substance abuse. First, Barrick, Mount and Strauss (in press) found evidence that highly conscientious individuals set more difficult goals for themselves and strive to accomplish them. Barrick et al. (in press) used the relationship between better job performance and the higher goals that individuals set for themselves to explain why conscientiousness predicts job performance. They argued that highly conscientious individuals will set more difficult goals for themselves which translates into better job performance.

Further, Schmidt and Hunter (1992) noted that highly conscientious individuals will spend more time on task which will also contribute to better job performance. However, improved job performance usually also entails the absence of substance abuse (e.g., McDaniel, 1988; Normand et al., 1990). Thus, integrity tests that seem to be assessing conscientiousness (Ones, 1993) may also correlate with and predict substance abuse.

A second explanation lies in the social impulse control enunciated by Gough (1948). According to this explanation, substance abusers are likely to be individuals who have not learned the social skills and social norms necessary to function effectively in society. They are deviants who have very poor impulse control. From this perspective, it could be argued that scores on integrity tests should also correlate with measures of substance abuse.

Finally, Zuckerman (1983) and his colleagues have posited that individuals differ in their proclivity to seek sensations. Individual differences in sensation seeking may be reflected in differences in integrity test scores and therefore such scores may be related to substance abuse.

CHAPTER III

METHODS

A thorough search was conducted to locate all existing integrity test validities for predicting the criterion of substance abuse. The literature was also searched for reliability and range restriction data on integrity tests. All published empirical studies referenced in the published reviews of the literature (O'Bannon et al., 1989; Sackett et al., 1989; Sackett & Harris, 1984), the three other meta-analyses of integrity tests (Harris, undated; McDaniel & Jones, 1986, 1988), and those identified through a computerized search of psychology and management related journals, were obtained. O'Bannon et al. (1989), located forty three integrity tests in use in the United States. All the publishers and authors of the forty three tests were contacted by telephone or in writing requesting validity, reliability, and range restriction information on their tests. Of these 36 responded with research reports. In addition, we identified other integrity tests overlooked by O'Bannon et al. (1989); their publishers were also contacted. All unpublished and published technical reports reporting validities, reliabilities, or range restriction information were obtained from integrity test publishers and authors. Some integrity test authors and test

publishers responded to our request for validity information on their test by sending us computer printouts that had not been written up as technical reports. These were included in the database.

Still other integrity test publishers responded to our request by sending us raw data that had not been analyzed. In some instances, using the information supplied, we were able to calculate the phi correlation, and then correct it for dichotomization (Hunter & Schmidt, 1990a). These corrected correlations were used in the meta-analysis. Thus, our database includes both published and unpublished data. The list of integrity tests contributing criterion-related validity coefficients, reliabilities, or range restriction information to this meta-analysis is presented in Table 1.

Insert Table 1 about here

Some researchers have argued for the exclusion of unpublished studies in all meta-analyses based on misleading and erroneous arguments that such unpublished studies constitute poor quality data. (The converse argument maintains that published studies have a positive

bias that overstates the results. Taken together, these two arguments will lead to scientific nihilism [Hunter & Schmidt, 1990b, p.515].) The hypothesis of methodological inadequacy of unpublished studies (in comparison to published studies) has not been established in any research area. In fact, ample evidence exists to prove the comparability of findings of published and unpublished studies in many research areas (Hunter & Schmidt, 1990b, pp. 507-509).

Hunter and Schmidt (1990b, pp. 509-510) present a hypothetical example that illustrates how differences between published and unpublished studies examining the effectiveness of psychotherapy could have been due to statistical artifacts. Ones et al. (in press) found that the correlation between the reported validity of integrity tests and the dichotomous variable indicating published versus unpublished studies is negligible. In the literature on the validity of employment tests, impressive evidence has been accumulated which indicates that published and unpublished studies do not differ in the validities reported (Hunter & Schmidt, 1990b, pp. 507-509). For example, the data used by Pearlman, Schmidt, and Hunter (1980) was found to be very similar to the U.S. Department of Labor (GATB) data base used by Hunter (1983) and other large sample military data sets. Also the mean validities

in the Pearlman et al. (1980) data base are virtually identical to Ghiselli's (1966) reported medians. Further, the percent of nonsignificant studies in the Pearlman et al. 1980 data base perfectly matches the percent of nonsignificant published studies reported by Lent, Aurbach, and Levin (1971). Finally, the percentage of observed validities that were nonsignificant at the .05 level in the Pearlman et al. (1980) data base (56.1% of the 2,795 observed validities) is consistent with the estimate obtained by Schmidt, Hunter, and Urry (1976), that the average criterion-related validation study has statistical power no greater than .50. If selectivity or bias in reporting were operating many of the nonsignificant validities would have been omitted, and the percent significant should have been higher than 43.9%. On the other hand, if unpublished studies were of poorer quality, not meeting the standards of peer review, then there should have been more than 56% non-significant validities among the unpublished studies. Thus, there is ample evidence arguing for the equivalence of published and unpublished studies. The two data bases are often comparable. Therefore, we included both published and unpublished reports in our analyses.

Data Coded or Extracted from Primary Studies

An identification number was given to each study. When more than one sample was reported in a study, a sample within study identification number was given to each sample within that study. Samples were numbered consecutively starting with the number one. Thus, each record contains a study identification number, a (within study) sample identification number, the validity coefficient, the sample size, the criterion used, whether the criterion measure was based on self-reports or external records, whether the sample was comprised of students or applicants to a job or current employees, and whether the validity coefficient was based on a predictive or a concurrent validation strategy. Wherever possible, we also coded the complexity levels of the jobs included in the analyses and other demographic characteristics.

Overall, we located fifty validation studies. Of these fifty studies, 24 had used employees as samples, 16 had used student samples, and the remaining ten studies were based on applicant samples. All fifty studies employed the concurrent validation strategy. Forty eight of the fifty studies relied on admissions of substance abuse. There was one study conducted in a sample of 46 employees in a fire department that had used apprehension and conviction for

substance abuse as the criterion. The observed validity coefficient in that study was .44. One study provided inadequate information as to whether admissions or external measures were employed. The observed validity coefficient in that study was .62 and it was based on a sample of 320 job applicants. Forty-seven of the fifty studies were on overt tests.

The admissions criterion was measured using self-report questionnaires. Measures of admissions of drug abuse included questions on number and type of illegal drugs used, number of times one has become "high" from drug use, etc. Measures of admissions of alcohol abuse included questions on frequency of alcohol intoxication, number of drinks consumed on the job, number of drinks on work breaks and during lunch on workdays, number of alcohol-related problems, etc. The final score was the sum (sometimes weighted) of such admissions.

Twenty of the fifty studies were conducted in the mid west while four were conducted in the north western region of the United States. Thirteen of the fifty studies were conducted in supermarket or grocery stores or convenience stores or gas station employees. Seven of the fifty studies were done using security personnel as sample. One study was conducted in a fire department while another was

in a fast food chain. Twenty studies focused on alcohol consumption while the remaining thirty used drug abuse as the criterion.

Given this set of validity coefficients, we could test the moderating influence of samples (Students, employees, applicants) and scales used. We also test the validities of integrity tests separately for predicting drug abuse and alcohol abuse. That is, we investigate whether (a) integrity tests have substantial validity in predicting the criterion of substance abuse; (b) the validity of integrity tests differs between student, employee, and applicant populations; and (c) drug scales of integrity tests have higher validity for predicting drug abuse than other scales.

Intercoder agreement in summarizing or extracting information from the primary studies is a concern in meta-analyses. Haring et al. (1981) present empirical data indicating that intercoder agreement in meta-analyses is a function of the judgmental nature of the items coded. The Haring et al. (1981) review of meta-analyses found that eight of the nine items lowest in coder agreement were judgments (e.g., the quality of the study) as opposed to calculation based variables (e.g., effect sizes, number of subjects). Jackson (1980) and Hattie and Hansford (1982,

1984) also provide data which indicate that problems of intercoder agreement in meta-analyses are negligible for coding computation-based numerical variables. Finally, Whetzel and McDaniel (1988) found no evidence of any coder disagreements in validity generalization data bases. The intercoder agreement in this research was over 85% for all categories coded. Disagreements between the coders were resolved through discussion.

Psychometric Meta-Analyses

Data from the sources described in the previous section was cumulated by the methods of psychometric meta-analyses. Depending on the availability of information in the primary studies, we can either correct the observed correlations for the effects of statistical artifacts and cumulate the individually corrected correlations, or use artifact distributions to correct the observed distribution of correlations, or use a combination of individual corrections and artifact distributions.

Because the degree of split for dichotomization is usually given in the research reports, it was possible to correct the correlations individually for the attenuating effects of dichotomization. But to correct for the effects of artifacts such as unreliability and range restriction, where the information available is sporadic, recourse was

made to the use of artifact distributions. That is, a mixed meta-analysis was employed. In the first step, the correlations were corrected individually for the effects of dichotomization. In the second step, the partially corrected distribution obtained from the first step was corrected for unreliability and range restriction using artifact distributions (Hunter & Schmidt, 1990b, p.188).

In correcting for dichotomization, sample sizes for the corrected correlations were adjusted to avoid underestimating the sampling error variance. First, the uncorrected correlation and the study sample size were used to estimate the sampling error variance for the observed correlation. This value was corrected for the effects of the dichotomization correction, and this corrected sampling error variance was then used with the uncorrected correlation in the standard sampling error formula to solve for the adjusted sample size, which was entered into the meta-analysis computer program. This process results in the correct estimate of the sampling error variance of the corrected correlation in the meta-analysis.

After the correlations were corrected individually for dichotomization, artifact distribution meta-analysis was used to correct for unreliability and range restriction. In using artifact distributions for correcting two or more

artifacts we have the option to use either the interactive procedure which corrects the observed correlations for the effects of the various statistical artifacts simultaneously, or the noninteractive procedure which corrects the observed correlation for the effects of the statistical artifacts sequentially (one after another). Recent computer simulation studies (e.g., Law, Schmidt, & Hunter, 1992; Schmidt et al., 1993) have shown that among the methods of psychometric meta-analyses the interactive procedure used with certain refinements, such as nonlinear range restriction and mean observed correlation in the sampling error formula, is the most accurate one.

The use of the mean observed correlation in the sampling error formula provides a more accurate estimate of the sampling error variance (Hunter and Schmidt, in press). The sampling error variance formula requires a knowledge of the population correlation. In individual studies, the observed correlation is taken as an estimate of the population value (because nothing better is available). But meta-analysts can be more precise by using the mean observed correlation across studies. This value is a better estimate of the population correlation than the individual observed correlation, which is strongly affected by sampling error unless sample sizes are large.

The second refinement involves the use of a nonlinear range restriction correction formula in estimating the standard deviation of true validities. In artifact distribution based meta-analyses, the mean and standard deviation of the residual distribution (the distribution of observed correlations expected when sample sizes are infinite and reliability and range restriction values are held constant across studies at their mean values) are corrected for the mean value of the artifacts. This procedure would be accurate if the artifact corrections were linear (e.g., reliability corrections), because the correction is the same for every value of the correlation in the residual distribution. But the correction for range restriction is not linear; it is smaller for larger correlations and larger for smaller correlations. This results in an overestimation of the true standard deviation when the linear approximation is used. Computer simulation studies have shown that a new, nonlinear correction procedure is more accurate (Law, Schmidt, & Hunter, 1993). That new procedure was used in this study.

More details of the refinements can be found in Schmidt et al. (1993) where examples are also provided to illustrate application of the refinements.

In correcting for unreliability in the measures, the use of the correct form of reliability coefficient requires the specification of the nature of the error of measurement in the research domain of interest (Hunter & Schmidt, 1990b, pp. 123-125). Three sets of artifact distributions were compiled for this technical report: one distribution for the reliability of the integrity tests, one distribution for the reliability of the criterion variables, and one distribution for range restriction. Descriptive information on the artifact distributions are provided in Table 2.

Insert Table 2 about here

A total of 124 integrity test reliability values were obtained from the published literature and the test publishers. Of the 124, 68 were alpha coefficients (55%) and 47 were test-retest reliabilities over periods of time ranging from 1 to 1,825 days (mean = 111.4 days; sd = 379.7 days). The mean of the coefficient alphas was .81 (sd = .10) and the mean of the test-retest reliabilities was .95 (sd = .10). There were 9 reliabilities reported without stating the type of reliability. The ideal estimate of

reliability for purposes of this meta-analysis is coefficient alpha or the equivalent. However, test-retest reliability estimates over relatively short time periods provide reasonably close approximations to alpha coefficients. Further, in this case the means of the two reliability types were similar. The overall mean of the predictor reliability artifact distribution was .81 and the standard deviation was .11. The mean of the square roots of predictor reliabilities was .90 with a standard deviation of .06.

No correction for predictor unreliability was applied to the mean true validity because our interest was in estimating the operational validities of integrity tests for selection purposes. However, the observed variance of validities was corrected for variation in predictor unreliabilities in addition to variation in criterion unreliabilities, range restriction values, and sampling error. For comparison purposes, we provide the percent variance due to sampling error alone in our results.

To estimate the reliability of the criterion measures, we reviewed the literature on delinquency and criminology. Viswesvaran, Ones, and Schmidt (1992) examined the appropriateness of self-reports of counterproductive behaviors by examining the correlations between admissions

and external measures. In that study, Viswesvaran et al. (1992) compiled a reliability distribution for admissions of counterproductive behaviors. They found 17 values, of which 13 were coefficient alphas and four were test-retest reliabilities. The 13 coefficient alphas comprised the criterion reliability distribution. The average of the reliability distribution was .84 and the standard deviation was .10. The average of the square roots of the reliability estimates was .94 and the standard deviation was .07.

Because integrity tests are used to screen applicants, the validity calculated using an employee sample may be affected by restriction in range. A distribution of range restriction values was constructed from the studies contributing to the database. There were 75 studies which reported both the study sample standard deviation and the applicant group standard deviation. The range restriction ratio was calculated as the ratio of study to reference group standard deviations (s/S). In four studies, correlations were reported for both the applicant and the employee groups. From these four studies range restriction ratios were calculated by taking the ratio of the two correlations reported and solving for the range restriction value using the standard range restriction formula (Case II formula; Thorndike, 1949, p. 173). Overall there were 79

range restriction values included in the artifact distribution. The mean ratio of the restricted sample's standard deviation to the unrestricted sample's standard deviation used is .81 and the standard deviation is .19, which indicates that there is considerably less range restriction in this research domain than is the case for cognitive ability (Alexander, Carson, Alliger, & Cronshaw, 1989). Thus, range restriction corrections were much smaller in present research than in meta-analyses in the abilities domain. No range restriction corrections were applied to student samples.

The parameters of interest estimated from a meta-analysis are the true validity, the standard deviation of the true validity, and the 90% credibility value. From the observed distribution of validities, we estimate the distribution of true validities. There are four substantive inferences of interest here. First, we want to know the average validity coefficient across situations. This is captured in the mean true validity. Second, we want to know whether the validity coefficient will be positive across situations. To answer this question we examine the 90% credibility value. The 90% credibility value indicates that in 90% of the situations the validity coefficient will be higher than this value. As such, if the 90% credibility value is positive, one can conclude

that the instrument has a validity coefficient that is positive in over 90% of the situations. That is, validity generalizes.

The third substantive question involves an examination of the standard deviation of true score validities to examine the extent to which the validity varies across situations. In a meta-analysis, if the 90% credibility value is greater than zero, but there is a sizable variance in the validities after corrections, it can be concluded that validities are positive across situations (i.e., validity generalizes), although the actual magnitude may vary across settings. However, the remaining variability may also be due to uncorrected statistical artifacts as well as methodological differences between studies. A final possibility is truly situationally specific test validities and/or the operation of moderator variables. In sum, the 90% credibility value is used to judge whether the validities are positive across situations (i.e., validity generalizes), whereas the estimated standard deviation of true score validities is used to assess whether the estimated true validity is constant across situations.

Finally, to test for the moderating influence of a hypothesized moderator, the validity coefficients are grouped into subsets based on the hypothesized moderator.

Psychometric meta-analysis is then conducted within each subset. If the hypothesized moderator exists, it will be reflected in the following findings: (a) the mean true validity computed for each subset will vary across the subsets, and will vary from the mean true validity computed with the entire set of validities across subsets; and (b) the average standard deviation of true score validities in the subsets will be lower than the overall standard deviation across. The above two results are interrelated as the group means and variances in the ANOVA paradigm, and together they test the extent of the moderating influence of the hypothesized moderator.

CHAPTER IV

RESULTS

The results of the psychometric meta-analyses of integrity test validities for predicting substance abuse (both alcohol and drug) are presented in Table 3.

Insert Table 3 about here

Based on all fifty samples, the mean true validity is .26. This represents a substantial level of validity. Further the 90% credibility value of .10 implies that the true validity will be greater than .10 in more than 90% of the situations. These values are based on a total sample size of 25,594.

The standard deviation of the true score validities is low (.14) which suggests that perhaps alcohol and drug abuse can be conceptualized as manifestations of the same phenomenon of substance or chemical abuse. That is, one might hypothesize that the same personality characteristics might underlie both alcohol and drug abuse.

The separate mean true validities for student, employee, and applicant populations are also provided in Table 3. In a selection setting, the focal population of interest is the applicant population. Many researchers have argued (see Ones et al., 1993, for a summary) that conscious and /or unconscious response distortion will affect integrity test validities. In taking these tests applicants have the greatest incentive for response distortion, followed by employees and students, in that order. That is, to the extent integrity test validities are affected by response distortion, true validities based on applicant samples should be lower than true validities based on employee samples, which in turn should be lower than the true validities computed on student samples.

The results reported in Table 3 confirm this expected gradient. But, although response distortion seems to attenuate the validity of integrity tests, its effects do not destroy validity. Even in the applicant population the true validity was .22 and the 90% credibility value was .14. Although this level of validity is moderate, these values suggest that the use of integrity tests in employment selection will translate into reduced levels of substance abuse in the workplace.

It is of interest to note that most of the sample consisted of applicants (about 90%). This is significant because applicants to jobs are our focus of interest. However, it would have been better if the applicant validities had been predictive in nature. The reader will recall that all validities in this meta-analysis are concurrent. The criterion for applicants was admissions of drug abuse made **at the time they were applicants**. Use of this same criterion measure taken after participants had been on the job for some time would have given a better indication of predictive validity. Since in predictive studies there may be less response distortion on the (admissions) criterion measure, predictive validity estimates might be higher than the .22 obtained here.

Specifically, with admissions as the criterion measure, concurrent studies done on applicants may underestimate predictive validity computed on applicants. Concurrent studies done on applicants using admissions will strongly lend themselves to response distortion on the criterion measure, which in turn would bias validity estimate downward. Applicants for jobs have strong incentives to minimize admissions of previous illegal drug use. Present employees already have jobs, and in addition are usually told that their responses will be used for research

purposes only. So present employees have much less incentive for response distortion on the criterion.

Given these biases, the actual operational validity of integrity tests for predicting drug abuse is probably somewhere between the validity of .22 (estimated with applicant samples) and .36 (obtained from employee samples). This would be a value large enough to produce practically significant reductions in substance abuse on the job if integrity tests are used in hiring.

Next, we analyzed the results of integrity tests for predicting alcohol abuse alone. The results are summarized in Table 4.

Insert Table 4 about here

The overall estimated true validity across 20 samples involving 1,402 individuals was .45 and the 90% credibility value was .29. The corresponding values in the employee population were .34 and .34, respectively. All the observed variation in validities computed on employee samples were attributable to statistical artifacts. In the student population, the true validity was .31 and the 90%

credibility value was .31 (again all the observed variation were explained by variations in statistical artifacts across the samples). There was only one study using applicants as sample; in that study the observed validity coefficient was .62. Studies using employee samples and studies using student samples had similar levels of validity, implying that response distortion is not a serious problem in employee samples for the criterion of alcohol abuse. However, the key question is the extent to which there is response distortion among applicants; the data here are too thin to really answer this question.

The results of the integrity test validities for the criterion of drug abuse are summarized in Table 5.

Insert Table 5 about here

Across student, employee, and applicant populations there were thirty studies including 24,192 individuals. Across the thirty studies the true validity was .25 and the 90% credibility value was .10. The true validity was highest in student samples and lowest in applicant samples indicating that response distortion may be affecting the validities of integrity tests for predicting the criterion

of drug abuse. However, the same caveats apply here as in⁵⁷
the case of alcohol abuse (Table 4).

Given the likely downward bias in the mean true validity derived from concurrent studies done on applicants, the actual operational validity of integrity tests for predicting drug abuse is probably somewhere between the validity of .21 (estimated with applicant samples) and .38 (obtained from employee samples). For prediction of alcohol abuse, the figure corresponding to this .38 is .34. (No meta-analytic estimate of the value for applicant concurrent validity was possible for the criterion of alcohol abuse.) Hence, the operational validity of integrity tests for predicting the two types of substance abuse may be very similar. We would speculate that in both cases operational validity is around .30--a value large enough to produce practically significant reductions in substance abuse on the job if integrity tests are used in hiring.

Some integrity tests (e.g., London House PSI) have subscales that are designed specifically for the purpose of predicting drug abuse. These scales have items asking the applicants about their attitudes toward drug and excessive alcohol use. The premise behind these items seems to be that individuals abusing alcohol and drugs will be more

lenient and accepting of others' abuse. On some overt integrity tests, there are also direct questions about past drug and alcohol use. The lengths of these scales are usually comparable to honesty scales of integrity tests, so are the reliabilities. The meta-analyses results of the validity of drug scales for predicting alcohol and drug abuse are presented in Tables 6 through 8. In many instances data were not available to analyze the validity for student, employee, and applicant samples separately. Further, the sample sizes were small in many analyses precluding the inference of robust conclusions. The results inferred from Tables 6 to 8 have to be very tentative.

Insert Tables 6 to 8 about here

It appears that in all analyses drug scales of integrity tests are valid predictors of both alcohol and drug abuse, the purpose for which they were constructed. We also investigated whether the drug scales have higher validity than the scales developed for predicting other counterproductive behaviors. The results for other scales are summarized in Tables 9 to 11.

Insert Tables 9 to 11 about here

The validities of honesty scales for predicting alcohol and drug abuse are presented in Table 9. Honesty scales of integrity tests ask job applicants about their attitudes toward theft in the work place. Some overt tests also include theft admission items on their honesty scales. On the surface honesty scales are very different from drug scales because honesty scales concentrate on attitudes and sometimes admissions of theft, while drug scales concentrate on attitudes toward and in some instances admissions of drug and alcohol use. In our analyses we found that honesty scales predict drug and alcohol abuse at levels comparable to drug scales. This is likely because both attitudes toward theft and drug and alcohol use are both stem from same underlying personality variables such as conscientiousness, agreeableness, and emotional stability. The fact that honesty scales predict drug and alcohol abuse at a level comparable to drug scales constructed specifically for that purpose is significant because this is one important piece of evidence that theft may be a marker variable for other types of counterproductive behaviors.

Table 10 reports the meta-analysis results of the validities of Nonviolence scales of integrity tests for predicting drug and alcohol abuse. Nonviolence scales of integrity tests ask applicants about their attitudes toward violent behaviors at work (e.g. fist fights). Some nonviolence scales also include items of admissions of past violent acts in the work place. In our analyses we found that nonviolence scales predict drug and alcohol abuse at levels somewhat lower than drug scales. However, because the total N in the nonviolence analyses was small (N=390), the possibility of sampling error causing this finding cannot be ruled out. The fact that nonviolence scales have positive moderate validity for a criterion they were not designed to predict, drug and alcohol abuse, is remarkable and may indicate that nonviolence also stems from the same personality variables that drug scales and honesty scales of integrity tests.

Finally, Table 11 presents the meta-analytic results for the validity of honesty and nonviolence scales for predicting drug abuse and alcohol abuse, separately. The small total sample sizes and the small number of correlations included in these analyses raise the suspicion that unaccounted sampling error could affect our conclusions. From the results reported in Tables 6 to 11, we can conclude that drug scales, honesty scales and

nonviolence scales appear to have comparable validity for⁶¹ the criteria of drug and alcohol abuse. This suggests that there is a common construct that is tapped into by drug scales, honesty scales, and nonviolence scales that is important for predicting the criterion of drug and alcohol abuse. However, the number of studies and subjects in these meta-analyses is too small for definitive conclusions.

CHAPTER V

DISCUSSION

The review of the literature on the constructs assessed by integrity tests resulted in the conclusion that integrity tests primarily assess conscientiousness, agreeableness, and emotional stability. The review of potential causal mechanisms indicated that conscientiousness, agreeableness, and emotional stability may be correlated with substance abuse. Based on these two streams of evidence, we developed our first hypothesis that all integrity tests will have substantial validity for predicting the criterion of substance abuse. Across fifty studies and situations, Integrity tests were found to have substantial validity.

Estimated true validity was higher in student populations than in employee population, and the estimated true validity in the employee population was in turn higher than the estimated true validity in the applicant population. This gradient in estimated true validity across the three populations is consistent with the hypothesis that individuals comprising the three populations have different levels of motivation for response distortion. But the literature on response

distortion in integrity tests has focused solely on response distortion on the predictor side. This exclusive focus on the predictor side is justifiable if the criterion was externally measured. When admissions are used as the criterion, we need to examine the potential for different levels of motivation in the three populations for response distortion on the criterion. Response distortion on the predictor will not bias estimates operational validity. In a real setting, applicants to jobs will engage in some response distortion on the predictor. The question becomes whether response distortion destroys predictive validity; and our results are in the negative. On the other hand, response distortion on the criterion will bias estimates of operational validity downward. Further, response distortion in admissions criteria will be more pronounced when the concurrent validation strategy is employed with applicants. That is, concurrent validities reported here underestimate the operational predictive validity of integrity tests.

Arguments have also been made (see Martin & Terris, 1990 for a summary) that the base rate of substance abuse is not known in the general population, and as such, we cannot estimate the utility of integrity tests for reducing the levels of substance abuse in the workplace. But the absence of an established base rate has no relevance for the validity

of integrity tests. More importantly, it is argued that with low base rates there will be more classification errors when integrity tests are used than when they are not used. Base rate refers to the proportion of test takers in the referent population who are actually substance abusers by some criterion. The argument is that integrity test usage results in high false positive rates (that is rejection of applicants who would not abuse drugs if hired) because the associated base rates are low (US OTA, 1990). (Note that usage of the terms false positive and false negative in integrity testing is the reverse of the regular usage of these terms in personnel selection. In an integrity test setting, a false positive error is the rejection of an applicant who would be a non-user if hired, and a false negative error is the acceptance of an employee who is a substance abuser.) This argument is based on an untenable assumption that all applicants would be accepted if an integrity test were not used. The failure to use any valid selection predictor will result in a higher false positive rate than its use. As validity increases, both false positives and false negatives decline. Therefore, any improvement in validity of the selection process will reduce both the probability of rejecting a qualified applicant and the probability of accepting an unqualified one. Hence, no matter what the actual base rate is, the validity of integrity tests cannot

be challenged on the grounds of low base rates. However, the utility of integrity tests to the organization does depend on the base rate in the applicant pool in that the larger this base rate (up to 50%) is, the greater will be the utility, other things being equal.

Some limitations of the present study need to be pointed out. First, a fully hierarchical moderator analysis was not possible. In fact even the main effects of some moderators could not be tested in this technical report. Further, the number of existing studies is small in certain analyses to raise concerns about the stability of the estimates. This has implications for second order sampling error in meta-analyses (Hunter & Schmidt, 1990b, pp. 411-450). But even with this limitation, a meta-analytic review based on a reasonable conceptual or theoretical framework provides sounder conclusions than other approaches to understanding the data, including the traditional narrative review. Future research should explore the moderating influences of job complexity, test type, etc.

The meta-analysis reported here is also noteworthy in that most of the studies reporting criterion-related validities for integrity tests used real applicants to jobs. This is significant because applicants to jobs are our focus of interest. In many predictor domains,

researchers have generalized results from students and employees to applicants which leaves the question of generalizability to applicants unaddressed. That is not the case in our analyses. However, it would have been better if the applicant validities had been predictive in nature and used externally measured criterion (instead of admissions). We need more studies with predictive designs using external measures of the criterion. Future research should build on our findings and test the conceptual and theoretical basis for these tests. Testing alternate causal mechanisms for the observed validity is another avenue for future research which may lead to increased understanding and better theories of work behavior and human motivation.

REFERENCES

- Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), Advances in experimental and social psychology (Vol. 2, pp. 267-299). New York: Academic Press.
- Alexander, R. A., Carson, K. P., Alliger, G., & Cronshaw, S. F. (1989). Empirical distributions of range restricted SD_x in validity studies. Journal of Applied Psychology, 74, 253-258.
- American Society for Personnel Administration (1988, June). Most employers test new job candidates, ASPA survey shows. Resource, Arlington, VA: Author.
- Ash, P. (1989). The construct of employee theft proneness. Rosemont, IL: SRA/London House, Inc.
- Barrick, M. R., Mount, M. K., & Strauss, J. P. (in press). Conscientiousness and performance of sales representatives: A test of the mediating effects of goal setting. Journal of Applied Psychology.
- Berry, R. E., & Boland, J. P. (1977). The economic cost of alcohol abuse. New York: Free Press.

- Bies, R. J. (1987). The predicament of injustice: The management of moral outrage. In L. L. Cummings, & B. M. Staw (Eds.), Research in Organizational Behavior (Vol. 9, pp. 239-319). Greenwich, CT: JAI Press.
- Bies, R. J., & Shapiro, D. L. (1987). Interactional fairness judgments: The influence of causal accounts. Social Justice Research, 1, 199-213.
- Blank, D. L., & Fenton, J. W. (1989). Early employment testing for marijuana: Demographic and employee retention patterns. In S. W. Gust & J. M. Walsh (Eds.), Drugs in the workplace: Research and evaluation data (National Institute on Drug Abuse Research Monograph No. 91, pp. 139-150). Washington, DC: U.S. Government Printing Office.
- Blocklyn, P. L. (1988, February). Consensus: Preemployment testing. Personnel, 66 - 68.
- Bureau of National Affairs, Inc. (1988, May). Recruiting and selection procedures, personnel policy forum survey no. 146. Washington, DC: Author.
- Conoley, J. C., & Kramer, J. J. (Eds.). (1989). The Tenth Mental Measurements Yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Cormack, R. W., & Strand, A. L. (1970). Trustworthiness Attitude Survey. Oakbrook, IL: Personnel Systems Corporation.

Crant, J. M., & Bateman, T. S. (1989). A model of employee responses to drug testing programs. Employees Responsibilities and Rights Journal, 2, 173-190.

Crouch, D. J., Webb, D. O., Peterson, L. V., Butler, P. F., & Rollins, D. E. (1989). A critical evaluation of the Utah Power and Light Company's substance abuse program: Absenteeism, accidents and costs. In S. W. Gust, & J. M. Walsh (Eds.), Drugs in the workplace: Research and evaluation data (Research Monograph 91, pp. 169-193). Washington DC: National Institute on Drug Abuse.

Cunningham, M. R. & Ash, P. (1988). The structure of honest: Factor analysis of the Reid Report. Journal of Business and Psychology, 3(1), pp. 54-66.

Freudenheim, M. (1988, December, 13). Workers' substance abuse increasing, survey says. New York Times, p.2.

Ghiselli, E. E. (1966). The validity of occupational aptitude tests. New York: Wiley.

Goldberg, L. R., Grenier, J. R., Guion, R. M., Sechrest, L.

B., & Wing, H. (1991). Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions. An APA task force report.

Washington DC.

Gomez-Mejia, L. R., & Balkin, D. B. (1987). Dimensions and characteristics of personnel manager perceptions of effective drug-testing programs. Personnel Psychology, 40, 745-763.

Gough, H. G. (1948). A Sociological theory of psychopathy. American Journal of Sociology, 53, 359-366.

Gough, H. G. (1954). Personnel Reaction Blank. Palo Alto, CA: Consulting Psychologists Press.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. Journal of Management, 16, 399-432.

Guastello, S. J., & Rieke, M. L. (1991). A review and critique of honesty test research. Behavioral Sciences and the Law, 9, 501-523.

Guthrie, J. P., & Olian, J. D. (1989, April). Drug and alcohol testing programs: The influence of organizational context and objectives. Paper presented at the Fourth Annual Conference of the Society of Industrial/Organizational Psychology, Boston.

Haring, M. J., Okun, M. A., Stock, W. A., Miller, W., Kinney, C., & Ceurvorst, W. R. (1981, April). Reliability issues in meta-analyses. Paper presented at the annual meeting of the American Education Research Association, Los Angeles.

Harman, H. H. (1976). Modern factor analysis. Chicago: The university of Chicago Press.

Harris, W. G. (Undated). An investigation of the Stanton Survey using a validity generalization model. Charlotte, NC: Stanton Corporation.

Harris, W. G. (1987). A components analysis of a pre-employment integrity measure: A replicated study. Psychological Reports, 60, 1051-1055.

Harris, M. H. & Sackett, P. R. (1987). A factor analysis and item response theory analysis of an employee honesty test. Journal of Business and Psychology, 2(2), pp. 122-135.

Hattie, J. A., & Hansford, B. C. (1982). Self-Measures and achievement: Comparing a traditional review of literature with a meta-analysis. Australian Journal of Education, 26, 71-75.

Hattie, J. A., & Hansford, B. C. (1984). Meta-Analysis: A reflection on problems. Australian Journal of Psychology, 36, 239-254.

Hay, D. W. (1981). An exploration of the dimensionality of a personnel selection instrument designed to measure honesty. Unpublished master's thesis, Illinois Institute of Technology, Chicago.

Herning, R. L., Glover, B. J., Koeppl, B. S., & Jaffe, J. H. (1989, September). Cocaine and workplace performance: Inferences from clinical studies. Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD.

Hill, M. F., Jr., & Sinicropi, A. V. (1987). Evidence in arbitration. Washington, DC: BNA Books.

Hogan, R. (1981). Hogan Personality Inventory. Minneapolis, MN: National Computer Systems Inc.

Hunter, J. E. (1983). Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the general aptitude test battery (GATB) (Test Research Report No. 45). Washington, DC: U.S. Employment Service, U. S. Department of Labor.

Hunter, J. E., & Schmidt, F. L. (1990a). Dichotomization of continuous variables: The implications for meta-analysis. Journal of Applied Psychology, 75, 334-349.

Hunter, J. E., & Schmidt, F. L. (1990b). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage

Jackson, G. B. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.

Jobs, S. (1989, September). Impact of moderate alcohol consumption on business decision making. Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD.

Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1986).

Use of licit and illicit drugs by America's high school students. Rockville, MD: National Institute of Drug Abuse.

Jones, J. W. & Terris, W. (1984). Validity and factor

structure of a dishonesty test that measures attitudes toward theft. (Tech. rep. #20). Park Ridge, IL: London House.

Klump, C. S. (1964). Stanton Survey. Charlotte, NC: Stanton Corporation.

Konovsky, M. A., & Cropanzano, R. (1991). Perceived fairness of employee drug testing as a predictor of employee attitudes and job performance. Journal of Applied Psychology, 76, 698-707.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1993).

Nonlinearity of range restriction corrections in meta-analysis: A test of an improved procedure. Manuscript submitted for publication.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971).

Predictors, criteria, and significant results. Personnel Psychology, 24, 519-533.

London House Inc. (1975). Personnel Selection Inventory.

Park Ridge, IL: Author.

London House Press. (1982). The Employee Attitude Inventory.

Park Ridge, IL: London House Press.

Lousig-Nont, G. M. (1987). Phase II Profile. Las Vegas, NV:

Author.

Martelli, T. A. (1988). Preemployment screening for honesty:

The construct validity, criterion related validity and test retest reliability of a written integrity test. Unpublished

Doctoral Dissertation, Ohio University.

McDaniel, M. A. (1988). Does pre-employment drug use predict on-the-job suitability? Personnel Psychology, 41, 717-729.

McDaniel, M. A., & Jones, J. W. (1986). A meta-analysis of the employee attitude inventory theft scales. Journal of Business and Psychology, 2, 327-345.

McDaniel, M. A., & Jones, J. W. (1988). Predicting employee theft: A quantitative review of the validity of a standardized measure of dishonesty. Journal of Business and Psychology, 2, 327-345.

Miller, J. F., & Bradley, P. (1975). Milby Profile.

Minneapolis, MN: Milby Systems Inc.

- Miller, J. D., Cisin, I. H., Gardner-Keaton, H., Harrell, A. V., Wirtz, P. W., Abelson, H. I., & Fishburne, P. M. (1983). National Surveys on drug abuse: Main findings. Rockville, MD: National Institute of Drug Abuse.
- Murphy, K. R., Thornton, G. C., & Reynolds, D. H. (1990). College students' attitudes toward employee drug testing procedures. Personnel Psychology, 43, 615-631.
- Normand, J., Salyards, S. D., & Mahony, J. J. (1990). An evaluation of preemployment drug testing. Journal of Applied Psychology, 75, 629-639.
- Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw Hill.
- O'Bannon, R. M., Goldinger L. A., & Appleby, G. S. (1989). Honesty and Integrity Testing. Atlanta, GA: Applied Information Resources.
- Ones, D. S. (1993). The construct validity of integrity tests. Unpublished Doctoral Dissertation, University of Iowa City, Iowa.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993).

Meta-Analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. Journal of Applied Psychology, 78 (August).

Paajanen, G. (1985). PDI Employment Inventory. Minneapolis, MN: Personnel Decisions, Inc.

Paajanen, G. E. (1987). The prediction of counterproductive behavior by individual and organizational variables. Unpublished doctoral dissertation, University of Minnesota.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980).

Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.

Reid Psychological Systems. (1951). Reid Report. Chicago, IL: Author.

Sackett, P. R., & Decker, P. J. (1979). Detection of deception in the employment context: A review and critical analysis. Personnel Psychology, 32, 487-506.

Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique.

Personnel Psychology, 37, 221-246.

Sackett, P. R., & Harris, M. M. (1985). Honesty testing for personnel selection: A review and critique. In H.

J. Bernardin & D. A. Bownas (Eds.), Personality assessment in organizations. New York: Praeger.

Sackett, P. R., Burris, L. R., & Callahan, C. (1989).

Integrity testing for personnel selection: An update.

Personnel Psychology, 42, 491-529.

Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance.

Current Directions in Psychological Science, 1, 89-92.

Schmidt, F. L., Hunter, J. E., & Urry, J. E. (1976).

Statistical power in criterion-related validation studies. Journal of Applied Psychology, 61, 473-485.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R.,

Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. Journal of Applied Psychology, 78, 3-12.

Science Research Associates. (1983). Personal Outlook Inventory. Parkridge, IL: Author.

Simpson, D. D., Curtis, B., & Butler, M. C. (1975). Description of drug users in treatment: 1971-1972 DARP admissions. American Journal of Drug and Alcohol Abuse, 2(1), 15-28.

Stecher, M., & Rosse, J. (1992, April). Attitudes toward random drug testing in the vaiation industry. Paper presented at the annual meetings of the Society for Industrial and Organizational Psychology, Montreal, Canada.

Stone, D. L., & Bommer, W. (1990, August). Effects of drug testing selection method and justification provided for the test on reactions to drug testing. Paper presented at the annual meeting of the Academy of Management, San Francisco, CA.

Stone, D. L., & Kotch, D. A. (1989). Individuals' attitudes toward organizational drug testing policies and practices. Journal of Applied Psychology, 74, 518-521.

Stone, D. L., & Vine, P. L. (1989, April). Some procedural determinants of attitudes toward drug testing. Paper presented at the Fourth Annual Conference of the Society of Industrial/Organizational Psychology, Boston.

Streufert, S., Pogash, R., Campbell, L., Kantner, A., Gingrich, D., & Jacques, C. H. (1991, January). Effects of alcohol consumption upon performance of white collar tasks. Paper presented at the National Institute on Drug Abuse Conference on Drug Abuse Research and Practice: An Alliance for the 21st Century, Washington, DC.

Thorndike, R. L. (1949). Personnel selection. NY: Wiley.

Trice, H. M., & Roman, P. M. (1972). Spirits and demons at work: Alcohol and other drugs on the job. Ithaca, NY: Hoffman Printing.

US Congressional Office of Technology Assessment. (1990). The use of integrity tests for pre-employment screening, OTA-SET-442. Washington, DC: US Government Printing Office.

Viswesvaran, C. (1993). Modeling job performance: Is there a general factor? Unpublished Doctoral Dissertation, University of Iowa City, Iowa.

- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1992, May). Appropriateness of self-reported criteria for integrity test validation: Evidence from beyond. In F. L. Schmidt (Chair), Criterion-Related validity, construct validity, and legality of integrity tests. Symposium conducted at the seventh annual meeting of the Society of Industrial and Organizational Psychologists, Montreal, Canada.
- Whetzel, D. L., & McDaniel, M. A. (1988). Reliability of validity generalization data bases. Psychological Reports, 63, 131-134.
- Winkler, H., & Sheridan, J. (1989, September). An examination of behavior related to drug use in Georgia Power Company. Paper presented at the National Institute on Drug Abuse Conference on Drugs in the Workplace: Research and Evaluation Data, Bethesda, MD.
- Woolley, R. (1991). An examination of the construct and criterion-related validity of overt and personality oriented predictors of counterproductivity. Unpublished masters thesis, University of British Columbia, Vancouver, Canada.

Woolley, R. M., & Hakstian, A. R. (in press). An examination of the construct validity of personality-based and overt measures of integrity. Educational and Psychological Measurement.

Vesavage, J. A., Leirer, V. O., Denari, M., & Holister, L. E. (1985). Carry-Over effects of marijuana intoxication on aircraft pilot performance: A preliminary report. American Journal of Psychiatry, 142, 1325-1329.

Zuckerman, M. (Ed.) (1983). Biological basis of sensation-seeking, impulsivity, and anxiety. Hillsdale, NJ: Erlbaum Press.

Table 1

Tests Contributing Data to the Meta-Analyses

Test Name
1. Accutrac Evaluation System ^a
2. Applicant Review ^a
3. Compuscan ^{a, c}
4. Employee Attitude Inventory (London House) ^a
5. Employee Reliability Inventory ^a
6. Employment Productivity Index ^b
7. Hogan Personnel Selection Series (Reliability Scale) ^b
8. Integrity Interview ^a
9. Inwald Personality Inventory ^b
10. Orion Survey ^{a, c}
11. P.E.O.P.L.E. Survey ^a
12. Personnel Decisions Inc. Employment Inventory ^b
13. Personal Outlook Inventory ^b
14. Personnel Reaction Blank ^b
15. Personnel Selection Inventory (London House) ^a
16. Phase II Profile ^a
17. P.O.S. Preemployment Opinion Survey ^{a, c}
18. Preemployment Analysis Questionnaire ^a
19. Reid Report and Reid Survey ^a
20. Rely ^a
21. Safe-R ^{a, c}
22. Stanton Survey ^a
23. True Test ^a
24. Trustworthiness Attitude Survey; PSC Survey; Drug Attitudes/Alienation Index ^a
25. Wilkerson Preemployment Audit ^{a, c}

Note. The list of publishers and authors of these tests are available in O'Bannon et al. (1989).

^aCovert integrity test. ^bPersonality-Based integrity test.

^cNo validity data was reported, but the test contributed to the statistical artifact distributions.

Table 2

Descriptive Information on Statistical Artifact Distributions Used to Correct Validities

	No. of values	Mean	Standard deviation	Mean of the square roots of reliabilities	Standard deviation of the square roots of reliabilities
Integrity test reliabilities	124	.81	.11	.90	.06
Criterion reliabilities	13	.84	.10	.94	.07
U (for range restriction correction) ^c	79	.81	.19	-	-

^cU refers to the ratio of the selected group standard deviation to the reference group standard deviation.

Table 3

Meta-Analyses of the Validity of Integrity Tests for Predicting Substance (alcohol and drug) Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	25,594	50	.20	.1175	.0984	.26	.14	13.1	29.9	.10
2. Employee samples	1,131	24	.28	.1295	0	.36	0	100	100	.36
3. Applicant samples	22,091	10	.17	.0710	.0538	.22	.07	8.5	42.5	.13
4. Student samples	2,372	16	.45	.1440	.1266	.48	.14	20.8	22.7	.31

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 4

Meta-Analyses of the Validity of Integrity Tests for Predicting Alcohol Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	1,402	20	.35	.1638	.0966	.45	.14	41.2	63.0	.29
2. Employee samples	644	16	.27	.1128	0	.34	0	100	100	.34
3. Applicant samples	320	1	.62	-	-	-	-	-	-	-
4. Student samples	438	3	.29	.0125	0	.31	0	100	100	.31

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 5

Meta-Analyses of the Validity of Integrity Tests for Predicting Drug Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	24,192	30	.19	.1075	.0909	.25	.13	10.0	28.4	.10
2. Employee samples	487	8	.30	.1468	.0561	.38	.08	64.5	85.4	.29
3. Applicant samples	21,771	9	.16	.0456	.0097	.21	.01	18.9	95.5	.20
4. Student samples	1,934	13	.48	.1444	.1280	.51	.15	19.3	21.5	.34

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 6

Meta-Analyses of the Validity of Drug Scales for Predicting Substance (alcohol and drug) Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	1,505	22	.31	.0759	0	.40	0	100	100	.40
2. Employee samples	400	11	.35	.0633	0	.44	0	100	100	.44
3. Applicant samples	357	4	.36	.0796	0	.45	0	100	100	.45
4. Student samples	748	7	.28	.0541	0	.29	0	100	100	.29

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; Sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 7

Meta-Analyses of the Validity of Drug Scales for Predicting Alcohol Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	826	13	.31	.0602	0	.40	0	100	100	.40
2. Employee samples	388	10	.33	.0835	0	.42	0	100	100	.42
3. Applicant samples	0	0	-	-	-	-	-	-	-	-
4. Student samples	438	3	.29	.0125	0	.31	0	100	100	.31

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 8

Meta-Analyses of the Validity of Drug Scales for Predicting Drug Abuse

Analyses categories	Total N	K	rmean	SDr	s res	p	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	670	9	.32	.0912	0	.41	0	100	100	.41
2. Employee samples	48	2	.38	.0390	0	.48	0	100	100	.48
3. Applicant samples	312	3	.38	.0613	0	.48	0	100	100	.48
4. Student samples	310	4	.25	.0772	0	.27	0	100	100	.27

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; p = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 9

Meta-Analyses of the Validity of Honesty Scales for Predicting Substance (alcohol and drug)

Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDρ	%	%	90% CV
for										
1. All samples	885	10	.37	.1636	.1098	.47	.15	31.7	54.9	.30
2. Employee samples	377	6	.29	.1366	.0267	.38	.04	72.3	96.2	.33
3. Applicant samples	104	1	.19	-	-	-	-	-	-	-
4. Student samples	404	3	.49	.1106	.0864	.52	.10	35.0	38.9	.41

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Table 10

Meta-Analyses of the Validity of NonViolence Scales for Predicting Substance (alcohol and drug)Abuse

Analyses categories	Total N	K	rmean	SDr	s res	ρ	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. All samples	390	7	.25	.0860	0	.32	0	100	100	.32
2. Employee samples	167	4	.20	.0482	0	.26	0	100	100	.26
3. Applicant samples	104	1	.20	-	-	-	-	-	-	-
4. Student samples	119	2	.36	.0436	0	.39	0	100	100	.39

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; ρ = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

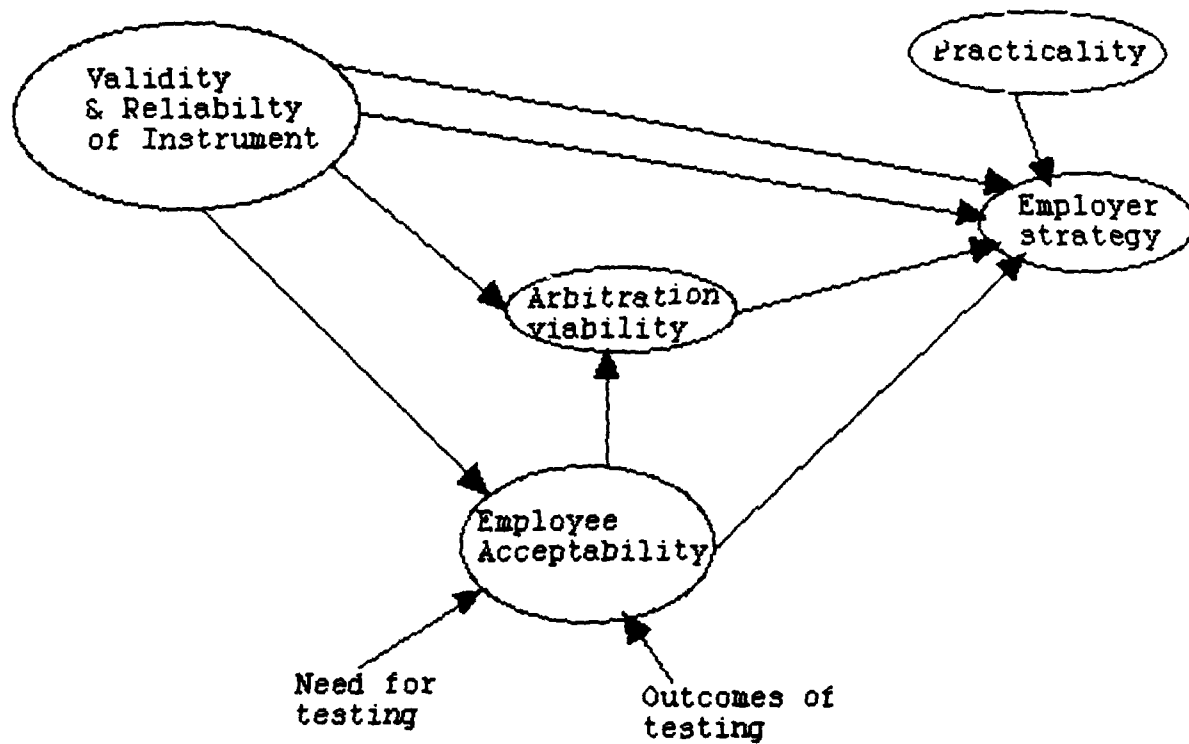
Table 11

Meta-Analyses of the Validity of Honesty and Nonviolence Scales for Predicting Alcohol and Drug Abuse Separately : Across All samples

Analyses categories	Total N	K	rmean	SDr	s res	p	SDp	%	%	90% CV
Var.S.E Var.acc for										
1. Nonviolence Scales predicting alcohol abuse	128	3	.19	.0545	0	.25	0	100	100	.25
2. Nonviolence scales predicting drug abuse	262	4	.28	.0856	0	.36	0	100	100	.36
3. Honesty Scales predicting alcohol abuse	128	3	.15	.1033	0	.20	0	100	100	.20
4. Honesty scales predicting drug abuse	757	7	.41	.1408	.0800	.52	.11	32.6	67.7	.39

Note. K = number of correlations; rmean = mean observed correlation; SDr = observed standard deviation; sres = residual standard deviation; p = true validity; SDp = true score standard deviation; % Var. S.E. = % variance due to sampling error; % Var. acc. for = % variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

Figure 1. Schematic Representation of Factors Affecting Employer Strategies



Grant Related Presentations and Publications

- Ones, D.S., Schmidt, F.L., & Viswesvaran, C. (May, 1993).
The nomological net for integrity tests. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Ones, D.S., Schmidt, F.L., & Viswesvaran, C. (May, 1993).
Integrity and ability: Implications for adverse impact and incremental validity. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Francisco, CA.
- Ones, D.S., Schmidt, F.L., & Viswesvaran, C. (June, 1993).
Meta-Analysis of integrity test validities: Implications for personnel selection and theories of job performance. Invited presentation at the annual conference of the International Personnel Management and Assessment Council, San Diego, CA. (The best student paper award).
- Ones, D.S., Schmidt, F.L., & Viswesvaran, C. (May, 1993).
Integrity tests predict substance abuse and aggressive behaviors at work. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada.
- Ones, D.S., Schmidt, F.L., & Viswesvaran, C. (August 1993).
Meta-Analysis of integrity test validities:
Implications for personnel selection and theories of job performance, Journal of Applied Psychology.

Schmidt, F.L., Ones, D.S., & Viswesvaran, C. (June 1993).
Integrity tests predict accidents on the job. Paper
presented at the annual meeting of the American
Psychological Society, Chicago, IL.

Schmidt, F.L., Ones, D.S., & Viswesvaran, C. (July 1993).
The validity of integrity tests for predicting substance
abuse. Paper presented at the NIDA conference,
Washington, D.C.

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (May 1993). Are
different aspects of job performance correlated? A
meta-analysis. Paper presented at the annual meeting of
the Society for Industrial and Organizational
Psychology, San Francisco, CA.

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (May 1993).
Theoretical implications of a general factor in job
performance. Paper presented at the annual meeting of
the Society for Industrial and Organizational
Psychology, San Francisco, CA.

Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (June 1993).
Integrity tests predict property damage at work. Paper
presented at the annual meeting of the American
Psychological Society, Chicago, IL.

Also coverage of our research on integrity in Fortune
magazine's March 8, 1993 issue and Redbook magazine's
August 1993 issue.